Validation

MUSIC® for Medical Students: Confirming the Reliability and Validity of A Multi-Factorial Measure of Academic Motivation for Medical Education

Tehmina Gladman[a], Steve Gallagher[b], Anthony Ali[c]

[a]Dean's Department, University of Otago, Wellington, New Zealand; [b]Dean's Department, University of Otago Dunedin School of Medicine, New Zealand; [c]Dean's Department, University of Otago, Christchurch, New Zealand

Key words: academic motivation, medical students, MUSIC, psychometric analysis, evaluation

Contact: Tehmina Gladman, Ph.D., tehmina.gladman@otago.ac.nz University of Otago, Wellington, PO Box 7343, Newtown, Wellington 6242, New Zealand. +64 4 918 6749.

Notes on contributors

Tehmina Gladman (PhD) is a lecturer and Education Adviser at University of Otago, Wellington. Her background is in experimental psychology and her research interests include academic motivation and motivation to use technology for learning.

Steve Gallagher (PhD) is a lecturer in eLearning at the Dunedin School of Medicine, with a background in behavioral psychology and an interest in learning technology and self-efficacy.

Anthony Ali is a professional practice fellow at the University of Otago, Christchurch with a background in teaching and learning. He has experience and interest in student assessment, course evaluation, and teacher professional development.

MUSIC® for medical students: Confirming the reliability and validity of a multi-factorial measure of academic motivation for medical education

Abstract

Construct: The MUSIC® Inventory measures the construct of academic motivation across five factors: empowerment, usefulness, success, interest, and caring. The factors are defined in terms of the degree the student perceives that they have control over their environment, that the coursework is useful to their future, that they can succeed in the course, that the course and instructional methods are interesting, and that the teacher cares about their wellbeing and their success respectively.

Background: A valid measure of medical students' academic motivation would provide medical teachers with a method for evaluating the motivational aspect of their course, and provide focus for changes in teaching and learning to improve medical student engagement. While the MUSIC® Inventory structure has been validated in the tertiary setting and with several professional programs, it has not been validated with medical students. The aim of this study was to use both classical test theory and Rasch modelling to assess the reliability and confirm the structure of the five-factor model of the MUSIC® Inventory with medical students.

Approach: 152 medical students completed the 26-item inventory. Descriptive statistics, internal consistency, correlations between factors, confirmatory factor analysis, and Rasch analysis using the rating scale model were performed to determine reliability and validity.

Findings: The five factors showed good internal consistency (Cronbach's alpha 0.87 – 0.92). Correlations between factors were moderate to high (r=.38 – .89). Confirmatory factor analysis highlighted inconsistencies in factor loadings of three of the items hypothesized to measure interest. Rasch analysis using the rating scale model showed that all items for each factor had good item fit (0.65 – 1.37). Person separation (2.28 – 2.85) and reliability (.84 –.91) scores indicated that the scales were able to differentiate different levels of respondents. Item separation (2.25 – 6.97) and reliability scores (.83 –.98) indicated that the items of the scales were being differentiated by the respondents.

Conclusions: Rasch analysis indicates that the five factors of academic motivation measured by the MUSIC® Inventory account for the response patterns in data from medical students. However, while the factors of empowerment, usefulness, success, and caring showed expected reliability and validity using classical analysis, three of the interest items cross-loaded on to the usefulness factor. Possible reasons may include ambiguity of language for the items or medical students' conception of usefulness and interest. Future research will explore medical students' understanding of the language used to measure these factors in further detail.

Keywords

academic motivation, medical students, MUSIC, psychometric analysis, evaluation

Introduction

Student engagement is a predictor of learning and performance.[1,2] The very act of being engaged adds to the foundation of skills essential to live a productive and satisfying life.[1] Students who are engaged in educationally productive activities enlarge their capacity for continuous learning and personal development.[3] The term student engagement refers to how involved or interested students appear to be in their learning and how connected they are to their classes, their institutions, and each other.[4] Reviews of the literature have classified engagement into three components: behavioral, emotional, and cognitive.[5,6] Several instruments have been designed to measure the level of student engagement at an educational institution and are increasingly seen as valid indicators of institutional excellence, including the National Survey of Student Engagement (NSSE) and the College Student Experiences Questionnaire (CSEQ).[7]

While these instruments have utility for general higher education, they tend to focus on observable aspects of engagement, such as length of involvement in certain activities over a certain time period and self-reported measures of interest.[8] Student engagement in health professional learning environments like medicine also requires cognitive engagement, where students are invested in their learning, embrace the challenge of acquiring new knowledge or skill sets, and go beyond the course requirements.[9,10] Health professional environments require students to "think deeply about the content and construct an understanding that entails integration and application of the key ideas of the discipline."[10 (p475)]

Medical education is no exception when it comes to conversations around how to measure and promote student engagement.[9] However, medical education requires students to be involved in their learning not only at a behavioral and emotional level, but also at a cognitive level if they are going to succeed in their subsequent practice. Students who have reached the primarily clinical phase of learning in medicine can encounter idiosyncratic learning situations that, more often than not, involve clinician teachers across varied discipline rotations.[11] Often, these clinician teachers have only sporadic and serendipitous interactions with students. This can make it challenging for the group of educators responsible for providing teaching and learning experiences in the medical degree to develop a sense of student engagement with the curriculum over time. This idiosyncratic environment is also challenging for students, and to remain cognitively engaged requires a level of academic motivation in the students.

Academic motivation

Academic motivation can be defined as "a process that is inferred from actions … and verbalizations … whereby goal-directed physical or mental activity is instigated and sustained."[12(p272)] Academic motivation is an important factor in learning,[13] increasing a learner's efficiency and helping them use their ability and talent to feel more satisfied.[14] The link between motivation, teaching, and learning has been extensively explored with several theories highlighting the multi-factorial nature of academic motivation.[15-17]

The relationship between engagement and motivation can be explored with reference to self-determination theory (SDT).[18,19] Simply stated, SDT posits that students will engage in their learning to satisfy their needs. Motivation is seen as a pre-requisite of and a necessary element for student engagement in learning.[19] Students who have high motivation make an effort to be engaged in learning opportunities;[20] therefore, educators can influence student outcomes by encouraging motivation.[21] A literature review of motivation in medical education identified evidence of the validity of SDT in this context, and factors that could be positively influenced by educators to improve motivation.[22]

Many clinician teachers and educators involved in the students' clinical learning phase could benefit from a simple, practical, and effective way to evaluate student engagement with their learning. By trying to determine the level of student cognitive engagement, a teacher can look at their current teaching context and instructional practices to determine if students are "simply participating or are willing to invest in learning and understanding".[10(p476)] One way to quantify a student's engagement in their learning is through measuring their academic motivation.

The Academic Motivation Scale,[23] based on self-determination theory,[18,24] relates students' learning autonomy and agency to their motivation to learn. While the Academic Motivation Scale has been extensively validated and used in research with medical students,[22,25,26] its focus on autonomy and agency overlooks other important aspects of academic motivation such as: utility value[27,28] and instrumentality;[29–31] self-efficacy,[32–34] expectancy-value theory,[28] and self-concept theory;[35] intrinsic interest value[27] and intrinsic motivation,[18] as well as flow;[36,37] and other concepts of relationships such as caring, belongingness, relatedness, and attachment.[38,39] To try to take better account of the many constructs theorized to influence academic motivation, Jones[17,40] developed the MUSIC® Model of Motivation (MUSIC® Model).

The MUSIC® Model

Jones[12,40] developed the MUSIC® Model to bring together the disparate theories and concepts of academic motivation into a clear and cohesive whole easily understood by teachers working with students of all ages,

> in any subject area (a) to design instruction that motivates students, (b) to diagnose strengths and weakness of instruction, and (c) to research relationships among factors critical to student motivation. [41(p3)]

The MUSIC® Model posits five factors that contribute to academic motivation: empowerment, usefulness, success, interest, and caring (Table 1).

TABLE 1 ABOUT HERE

From this theoretical model, Jones and Wilkins[42] developed and validated the MUSIC® Model of Academic Motivation Inventory (MUSIC® Inventory) with 26 items made up of 4 to 6 statements for each of the five factors in the MUSIC® Model. The instrument can be delivered at any time to evaluate an individual activity, a course, or a full program. Students rate each statement, presented in random order, with 6-point Likert-type items (from 1=strongly disagree to 6=strongly agree). Students typically take less than 10 minutes to complete the inventory. The instrument is easily scored by calculating the mean of the scores of the statements for each factor, with higher values representing higher levels of the factor. Importantly, each factor is analyzed separately, enabling teachers to focus on particular aspects of academic motivation. The inventory does not give an overall academic motivation score. Jones includes versions of the MUSIC® Inventory that may be used with students of all ages as well as an educator version.[41]

When testing the validity of a measure in a new population, a common method for doing so is to test its internal consistency and use confirmatory factor analysis to confirm the underlying structure of the measure. The College Student version of the MUSIC® Inventory was first validated in this way with university students in the U.S. which showed strong inter-item consistency while confirming a reasonable factor structure. [40,42] The inventory has also been validated with tertiary students in other countries such as Iceland,[43] China,[44] and Turkey[34,43] with similar results confirming the overall factor structure and internal consistency of the items.

The MUSIC® Inventory has also been validated with other university student groups in a number of professional courses. However, not all validation studies have found clear evidence of the theorized five-factor structure in the inventory. In a study of three samples of veterinary medicine students, a shortened 20-item MUSIC® Inventory was tested and found to have good psychometric properties, though one sample showed higher correlations between interest and usefulness items than previously noted. [45] A study of three cohorts of pharmacy students determined that the full MUSIC® Inventory showed good internal consistency, and confirmatory factor analysis showed reasonable underlying structure; however, two of the interest items had low factor loadings (<0.5) on the interest factor.[46] Tendhar, in a study of engineering students, found good internal consistency for the items measuring each factor; however, exploratory factor analysis found evidence of overlap in the factor loadings of interest and usefulness items. [47]

While findings for the MUSIC® Inventory across several university samples have shown a high degree of internal consistency and validity, testing of students in professional degree programs exposes potential limitations in the five-factor model. Validation studies have been conducted in the health professions of Pharmacy [46] and Veterinary Medicine;[45] however, in both of these cases the students who provided data for the validation studies appear to be learning in primarily non-clinical components of their course at the time the data were collected. Many health professional programs involve learning in clinical environments, and this is a particular feature of medical education.[22] Clinical learning emphasizes student observation and participation in clinical practice, and involves a large number of clinician educators who work within a hard to control environment of frequent changes as students move through clinical rotations and placements.[11,48] Factors such as the year of study in medical programs and teacher support have been identified as potential determinants of motivation in medical education.[22] For these reasons, it is important to validate any measures of academic motivation with students who are clearly participating in clinical learning environments.

The MUSIC® Model and its associated Inventory could provide medical teachers with a clear method for evaluating academic motivation in their courses; however, it is important that it first be validated with medical students to ensure that the structure remains robust in a primarily clinical education environment. As it has been successfully validated in a wide range of contexts, we hypothesized that the MUSIC® Inventory would be valid for use with medical students in New Zealand. As a result, our research question was: Is the five-factor model of the MUSIC® Inventory valid when used with medical students who have reached the primarily clinical learning phase of their instruction? The aim of this research was to complete reliability and validity testing to determine if the underlying factor structure of the MUSIC® Inventory is present in a medical student group.

Methods

Participants and Procedures

Participants were medical students in their Advanced Learning in Medicine (ALM) program (years 4-6) at the University of Otago Medical School (OMS) in New Zealand, a three-year clinical clerkship program that follows the two-year Early Learning in Medicine program. Students attend one of three campuses for ALM, where one campus has approximately 240 students and two have approximately 360 each.

Students were contacted via the medical school's learning management system and asked to complete the MUSIC® Model of Academic Motivation Inventory[41] online to measure their general perceptions of the medical program as whole, encompassing their perceptions of their pre-clinical and clinical learning from the perspective of a student in a clinical learning environment, and demographic questions. Participation was voluntary. Consent was obtained as an entry point to the questionnaire. The survey was open from 23 August to 30 September, 2018 (5.5 weeks) with two reminders. The online survey was set up according to detailed instructions in the user manual for the College Student, present tense, major/program level version.[41]

Instrument

Students completed the MUSIC® Model of Academic Motivation Inventory (College Student, present tense, major/program level version) which includes statements measuring empowerment (5 statements), usefulness (5 statements), success (4 statements), interest (6 statements) and caring (6 statements). Items were ranked on the following scale: 1=Strongly Disagree, 2=Disagree, 3=Somewhat Disagree, 4=Somewhat Agree, 5=Agree, 6=Strongly Agree (Appendix 1). Mean scores were then calculated for each set of items leading to five scores on the inventory for each student.[41] As recommended,[41] the inventory was adapted slightly to terminology used in the medical program. Four terms were changed to reflect local terminology: 'course(s)' was changed to 'module(s)', 'instructor(s)' was changed to 'tutor(s)', 'grade' was changed to 'mark', and 'instructional' was changed to 'teaching.' Students were also asked to complete six demographics questions asking them for their age range, gender (using standard New Zealand census language), campus, year (4th, 5th or 6th), ethnic group (New Zealand census choices) and mode of entry into the medical program (via first year health sciences, graduate, or alternative entry).

Analysis Methods

Descriptive statistics and classical test analysis, including confirmatory factor analysis, were completed using RStudio, Version 1.2.1578. [49] Rasch analysis was performed with Winsteps, Version 4.4.7.[50]

Descriptive Statistics

Demographic characteristics of the student sample were calculated. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was calculated (using the psych 1.8.4 package[51]) to determine if our sample size had sufficient power to conduct a confirmatory factor analysis. KMO values over .70 are considered to indicate adequate sampling for factor analysis.[52] Bartlett's test of sphericity was used to measure the suitability of the data for confirmatory factor analysis. Bartlett's test of sphericity compares the data correlation matrix to the identity matrix to determine if they are equivalent. If the correlation matrix equals the identity matrix then the variables have no relationship with each other and an analysis of structure would be inappropriate.[52]

Classical test analysis

Internal consistency of each of the five scales of the inventory was assessed by calculating Cronbach's[53] alpha values. Cronbach's alpha is a measure of the internal consistency of a scale and is one method for determining the interrelatedness of items within a scale.[54] Cronbach alpha scores of .70 or higher are considered to be a good indication of internal consistency of a scale.[54] Inter-scale correlations were tested using Pearson's correlation coefficient. Moderate positive inter-scale correlations were expected, based on previous studies.[40,45]

Confirmatory factor analysis

Confirmatory factor analysis (CFA) was conducted using lavaan 0.6-1 package in R[55] to determine the fit of the five-factor structure theorized in the MUSIC® Model in this student group. We also scrutinized the factor loadings of each item of the 26-item Music® Inventory to determine if the inventory was fit for purpose. As in previous work,[40,45] it was predicted that the individual items of each scale of the inventory would load on to their expected factor.

Rasch analysis

Rasch analysis was conducted using Winsteps, version 4.4.7[50] to consider the item and person fit for each factor of the MUSIC® Inventory to the model. Rasch analysis, based on the Rasch model, is a type of latent trait modelling that transforms ordinal data to interval data by converting the raw data to logits. Logits have the properties of interval scaling and allow for probabilistic modelling of data.[56] A Rasch analysis assumes that a scale has a single underlying dimension, or latent variable, and measures the relationship between actual responses to items on the scale and a theoretical model of the responses.[57] This measure allows us to determine how well the scale separates people based on their responses, and the distribution of the items measuring the latent variable of interest from those easiest to endorse by respondents to those most difficult to endorse.[58] This allows us to determine if the items for the scale represent the breadth of the latent variable and if there are redundant items in the scale.[57] It is generally considered appropriate to use Rasch analysis as an alternate psychometric method when analyzing the structure of an existing measure that uses ordinal data.[59]

The first step was to conduct a test of dimensionality to confirm that the MUSIC® Inventory was multidimensional and therefore that it was appropriate to treat each scale within the inventory as a separate unidimensional measure. We then tested the dimensionality of each individual scale to confirm that it represented a unidimensional measure. To test the fit of our data to the model, we used the rating scale model[60,61] to test item fit using information-weighted fit (infit), outlier-sensitive fit (outfit), and point-measure correlation for each scale.

The rating scale model was used because the MUSIC® Inventory uses common Likert-type items for each scale.[62] It is also more robust in contexts where items may have fewer than ten responses within some categories. Finally, the rating scale model uses a fixed threshold which makes the fewest assumptions about the data.[62,63] We tested scale reliability using both person separation and reliability, and item separation and reliability. Person separation indicates how well the items differentiate between people being measured[64,65] while item separation considers how well respondents are able to differentiate between items.[57,65] Both person and item reliability can be interpreted in a manner similar to Cronbach alpha values.[57]

Results

Descriptive statistics

In total 156 students out of 882 students across the three campuses completed the survey (152 useable for analysis), giving a 17% response rate. One student did not complete the demographic data questions. Most students were 21 to 24 years of age (108/155, 69.7%), female (92/155, 59.4%), NZ European (109/155, 70.3%), and had entered through the first-year health science category (109/155, 70.3%). There were 64 (41%) 4^{th} year students, 38 (25%) 5^{th} year students, and 53 (34%) 6^{th} year students. Students represented all three campuses with 62 students from Campus 1, 43 from Campus 2 and 50 from Campus 3. The demographics of the respondents were broadly representative of the population of ALM students (Table 2) with the exception of ethnicity. The sample was overrepresented by NZ Europeans ($\chi 2 = 10.94(3)$, $p = .012$).

TABLE 2 ABOUT HERE

The KMO was .91, which confirms adequate sampling. Bartlett's Test of Sphericity was significant ($\chi 2 = 3122.46$, $p < 0.001$), indicating that the correlation and identity matrices were not identical, therefore the data were adequately correlated to perform a structure analysis.

Classical test analysis

Inter-item correlations were in the high-moderate to high range for each of the expected scales (Cronbach's alpha .87 – .92) with scale intercorrelations (Pearson's correlation coefficients) showing expected moderate correlations between all but one pair of the five factors (Table 3).

TABLE 3 ABOUT HERE

The correlation between usefulness and interest was stronger than expected (*r* = .89) indicating possible overlap between these two factors.

Confirmatory factor analysis

Confirmatory factor analysis was calculated stipulating a five-factor solution (Figure 1) with oblique rotation. Fit indices showed a $\chi2$ of 633.170, comparative fit index (CFI) of .885, Tucker-Lewis index (TLI) of .870, root mean square error of approximation (RMSEA) of .089, and a standardized root mean square residual (SRMR) of .062. For standard cut-off criteria for measures of fit, we would expect to see a CFI of .90 or higher with RMSEA and SRMR of .05 or smaller.[66] As values for the five-factor solution did not show a good enough fit to consider the model to be properly identified,[66,67] a post-hoc scrutiny of the factor loadings of the items was undertaken.

Scrutiny of the factor loadings of the individual items showed that while items for four of the factors loaded as expected, three of the interest items loaded on to the usefulness factor (Table 4). The statements representing the interest construct that loaded onto the same factor as the usefulness items were those items focused on the coursework itself, while the interest items related to the instructional methods in the course loaded on the interest factor.

To try to determine if a more appropriate structure was present, we undertook a further set of factor analyses based on the path diagram and factor loadings. An adapted five-factor analysis was performed which collapsed the usefulness items with the three interest items that loaded onto the same factor with usefulness, leaving the three further interest items as their own factor. Fit indices were similar to the theoretical five-factor model ($\chi2$ = 603.17, CFI = .89, TLI = .88, RMSEA = .085, SRMR = .061). A six-factor model also showed similar fit indices ($\chi2$ = 575.04, CFI = .90, TLI = .89, RMSEA = .082, SRMR = .059). Finally, we performed an adapted five-factor model without the three poorly loading interest items to determine if removal of these three items would improve the fit of the structure. Fit indices showed greatest improvement with the three coursework interest items removed ($\chi2$ = 467.17, CFI = .91, TLI = .89, RMSEA = .086, and SRMR = .061). Consideration of the factor loadings of the individual items also shows stronger factor loading for items overall when the three interest items were removed from analysis (Table 4).

TABLE 4 ABOUT HERE

Rasch Analysis

Overall item results

To confirm that the MUSIC® Inventory is multidimensional, a test of dimensionality was completed on the full inventory. The test of dimensionality exposed three contrasts with eigenvalues over 2 (2.70 – 4.83) and two further contrasts with eigenvalues of 1.43 and 1.89. As contrasts with eigenvalues over 2 are generally considered to represent sets of items that are different from the noise of unexplained variance in the model, at least three separate factors can be said to exist in the dataset.[50]

Table 5 shows the item infit and outfit mean squares for each item in the scales, person separation and reliability for each scale, and item separation and reliability for each scale. Infit and Outfit mean squares for all items in each of the five scales was between 0.65 and 1.37, with positive point-measure correlations, indicating a good fit of the items to the model and no disordered response categories.[68] The person reliability for each of the five factors indicated high reliability for the items (.84 – .91).[64] The person separation for each of the scales was also good, showing that the items were able to separate respondents appropriately (2.28 – 2.85).[58] Item separation was also within acceptable limits (2.25 – 6.97)[58] with similar item reliability (.83 – .98).

TABLE 5 ABOUT HERE

Scale Results

To ensure that each factor was unidimensional, a test of dimensionality was performed. For four of the five factors – empowerment, usefulness, success, and caring – no contrast was found to measure over 2.0, suggesting that those factors are unidimensional.[39] Table 5 shows the item difficulties from most difficult (highest score) to easiest (lowest score) for the items that make up each of the scales of the MUSIC® Inventory. Item difficulties can be thought of as the ease with which a respondent will endorse the item. For example, medical students found the item, "I have flexibility in what I am allowed to do in the courses." the most difficult empowerment item to endorse or agree with, while the item, "I have control over how I learn the course content." was the easiest empowerment item to endorse or agree with. When testing the Interest items for dimensionality, one contrast showed an eigenvalue of 2.15, indicating a possible second dimension in the item set. Items 2–4 were located within cluster one of contrast one and items 1, 5, and 6 were located within cluster three of contrast one. This clustering is similar to the factor loading differences found in the confirmatory factor analysis.

Person-item maps (Wright Maps) graphically present the distribution of person responses to items on the left side of the map, and the distribution of item difficulties on the right side of the map. The mean person distribution is represented by the M on the left of the figure, and the mean for the item difficulties is represented by the M on the right of the figure. The Wright Maps for each of the scales are presented in Appendix 2. For the empowerment and interest scales, the person distribution and item difficulty means are less than 1 logit apart indicating that the two distributions are reasonably well aligned.[53] However, usefulness, success and caring show differences between the means of over 1.5 logits indicating a possible ceiling effect in the endorsement of items. Several items of the empowerment, usefulness, and success scale cluster within two logits of the mean, indicating possible redundancy in the items (Appendix 2). The interest scale items show that items 1 ("The coursework holds my attention") and 6 ("The coursework is interesting to me") of the interest scale appear to be consistently easier to endorse than items 2–5 (Appendix 2). The items also line up together on two points in the distribution, with items 2–5 approximately one logit above the mean, and items 1 and 6 nearly two logits below the mean. From the map, the items appear to be redundantly measuring two aspects of interest. Of all the scales, the items in the caring scale show the broadest distribution, though there is a significant gap between items 3 and 6 at the more difficult end of the map, and items 1, 2, 4, and 5 at the easier to endorse end of the map. There is no redundancy in the items, but the gap indicates the possibility that there is an opportunity to improve the scale with the thoughtful addition of items to fill the space between the easy to endorse and more difficult to endorse items.

Appendix 2 includes the probability curves for each response category for the five scales. From these graphs, each category shows a clear area, or threshold, along the continuum where it is more likely to be chosen. This confirms the Likert-type items are differentiating people at each category threshold with no disordering of the categories.

Discussion

For medical students in later stages of their course, the MUSIC® Model of Academic Motivation Inventory can be used to measure the five individual scales in the model with high degrees of internal consistency and reliability. As a first step towards validating the model in medical students who are in the clinical phase of instruction, this is a promising measure of its respective underlying constructs of academic motivation. Very good internal reliability of each of the five individual scales was seen with Cronbach's Alpha of .87 – .92, which confirms previous research showing that the five individual scales are internally consistent and reliable.[40,46] However, the inter-scale correlations were higher than seen in previous validity research for the MUSIC® Inventory,[40,45,46] indicating possible overlap of scale concepts in this student context. If a correlation of .69 is the cutoff for moderate correlations (meaning the scales share 48% of the variance),[69] then the interest scale is highly correlated with usefulness (r = .89).

The confirmatory factor analysis also indicated a somewhat weaker fit for the five-factor model than seen previously.[40,45,46] Although competing models were tested, including a five-factor model based on the factor loadings and a six-factor model separating the two aspects of the interest factor, none showed an improvement in model fit as compared to the theoretically constructed five-factor model. From this we may speculate that a possible higher order or bifactor model of the interest data may be involved. However, while our data are consistent with that found in the studies of pharmacy and veterinarian students,[45,46] before considering alternative models, it would be appropriate to first establish whether these results are replicated in similar samples drawn from different medical student populations.

The Rasch analysis indicated that overall the items fit well in the model, with infit and outfit mean squares in the range considered to be well-specified.[68] Based on previous research, person and item separation scores over 2.0 may be considered good differentiation while scores over 3.0 show excellent differentiation.[57,58] As separation indices for both person and item separation in this sample were over 2.0, the individual scales of the MUSIC® Inventory were able to differentiate across the range of students completing the inventory as well as across the range of items within the scales. However, several of the scales showed slight ceiling effects in the person distribution and some redundancy in the items. Our data suggest that the items used to measure each of these scales could be refined; however, this was outside the scope of the current research.

Possible explanations for the anomalous results found in the confirmatory factor analysis may relate to differences between this New Zealand medical school student cohort and previously researched student groups. For example, these medical students might be misinterpreting the language used in the items of the MUSIC® Inventory. While some items were appropriately adjusted to reflect local education language preferences, perhaps the language itself was ambiguous. Since these students were in a clinical clerkship period of their course, they may find the term “coursework” confusing. We wanted them to interpret coursework to mean both in-class types of assignments as well as the clinical and independent learning opportunities, but this may have been too broad a definition.

It is also possible that the variation across the three clinical years and three campuses has added variance to the results of the program version of the inventory. Previous studies focused on validating the individual course version of the inventory rather than the program version.[40,45,46] A previous validation study with students in a veterinary medicine programme[45] combined data from different years of study within the program in order to create appropriately sized samples for analysis, and showed that one sample had higher correlations between two MUSIC scales than the others. This suggests that differences in learning context may influence the stability of the factor structure.

Finally, the poorer fit of the interest items in our model with our dataset could be due to how our medical students are conceptualizing the items measuring interest. This inference has some support from the recent work by Tendhar[47] who found an overlap between interest and usefulness in his work with engineering students; the study of pharmacy students by Pace and colleagues[46] which found that the same three interest items showed lower loadings on the interest factor; and the work by Jones, Byrnes and Jones[45] which found high correlations between usefulness and interest (r = .79) in one of their samples of veterinary school students. In each of these three cases, the students are part of a program where they already have a strong sense of professional identity. In these cases, the students may consider useful coursework to be interesting by definition, particularly if the specific questions could be misconstrued as being related to personal interest rather than the model focus of situational interest (Table 1).

Overall, the results show promise for the use of the MUSIC® inventory in measuring academic motivation in medical students. The inconsistencies with some previous work demonstrate the importance of validating instruments with particular student groups whose educational context differs from previously validated student groups. In addition, the pattern of inconsistencies found in this study and other work[45–47] in professional programs suggests that further research to tease out the apparent overlap between interest and usefulness in professional learning contexts is justified. Despite these inconsistencies, the factor structure is generally supported by our analysis and we consider that the MUSIC® inventory can be used in medical education contexts to support educational and professional development where issues with student motivation are identified. For example, using the inventory to assess motivation within different components of a medical program might identify areas where motivation is lower than other parts of the course. In this situation, the medical school could investigate whether curriculum modification or teacher professional development is needed to address issues with student academic motivation. Previous research with pharmacy students, for example, identified part of the course with lower expectations for student success by using the MUSIC® inventory, and used this information to implement an intervention to address this aspect of academic motivation.[46]

Limitations

Although this study was undertaken with care, there are some limitations that should be acknowledged when interpreting the data. Despite the KMO analysis confirming that the sample size was adequate for the validity study, the low response rate and lack of ethnic representation in the sample introduced the possibility of non-response bias in our results. The possibility exists that students who responded are more similar to one another than students who did not respond which could have impacted the correlation and CFA results. It might be that our sample is biased to include more highly motivated students, which might have the effect of increasing our scores on the MUSIC scales. There is some evidence of ceiling effects as visualized in the Wright maps**,** and a biased sample is one possible explanation for this effect. However, further work with a higher response rate would be required to investigate this possibility. We also obtained data from a relatively representative sample in terms of campuses, gender, and year (when compared to the most recent publicly available data[70,71]) giving us some confidence in our results. Instrument items were not pre-tested with a pilot sample to ensure that the terms used were understood similarly by this student population as the item language had been adapted to the local context. We intend to address these limitations in future work.

Future Work

Our next step will be to engage with a sample of students using qualitative methods to determine student perceptions of the language used in the program version of the MUSIC® Inventory. At this time, we may be able to tease out whether students are conceptualizing the interest items as defining personal interest or situational interest. Alongside this work, we will attempt to replicate our results with a larger and more diverse sample of medical students to develop a clearer picture of how medical students conceptualize usefulness and interest.

To determine if the sample size or the program version of the inventory influenced our results, we plan to conduct discriminant and convergent validity testing using the present and past tense College Student version of the MUSIC® Inventory with a larger sample size, focusing on specific modules and within specific year groups. We also plan to begin the process of validating the MUSIC® Inventory for staff in order to compare staff and student perceptions of the academically motivational aspects of a module in medical education.

Conclusion

The aim of this research was to complete reliability and validity testing to determine if the underlying factor structure of the MUSIC® Inventory is present in a medical student group. While our initial evaluation of the full MUSIC® Inventory suggests the program version of the instrument incorporating the full interest scale is not yet entirely appropriate for use with medical students, the modified MUSIC® Inventory appears to be a reliable and internally valid measure of its respective underlying constructs of academic motivation. With the future work detailed above, the MUSIC® Inventory has potential to give medical teachers a clear measure of students' academic motivation, enabling them to determine where the strengths and weaknesses lie in their current context and strategically design teaching and learning opportunities that increase student engagement.

Tehmina Gladman https://orcid.org/0000-0002-5112-3460
Steve Gallagher https://orcid.org/0000-0003-1108-4388
Anthony Ali https://orcid.org/0000-0002-8835-2937

Acknowledgements

We would like to thank Associate Professor Rebecca Grainger for giving us thoughtful feedback as internal peer reviewer for this article. We would also like to thank Dr. James Green for his assistance in the Rasch analysis of the data.

Aspects of this work were presented at the 2019 Australian and New Zealand Association for Health Professional Educators (ANZAHPE) and the 2020 Ottawa Conference on Assessment.

Ethical Approval

Ethical approval was granted by the University of Otago Human Ethics committee (Ref number D18/048).

Declaration of interest

The authors report no declarations of interest.

References

1. Carini RM, Kuh GD, Klein SP. Student engagement and student learning: Testing the linkages*. *Res High Educ*. 2006;47(1):1-32. doi:10.1007/s11162-005-8150-9

2. Pascarella ET, Seifert TA, Blaich C. How effective are the NSSE benchmarks in predicting important educational outcomes? *Change Mag High Learn*. 2010;42(1):16-22. doi:10.1080/00091380903449060

3. Shulman LS. Making differences: A table of learning. *Change Mag High Learn*. 2002;34(6):36-44. doi:10.1080/00091380209605567

4. Heaslip G, Donovan P, Cullen JG. Student response systems and learner engagement in large classes. *Act Learn High Educ*. 2014;15(1):11-24. doi:10.1177/1469787413514648

5. Fredricks JA, Blumenfeld PC, Paris AH. School engagement: Potential of the concept, state of the evidence. *Rev Educ Res*. 2004;74(1):59-109. doi:10.3102/00346543074001059

6. Jimerson SR, Campos E, Greif JL. Toward an understanding of definitions and measures of school engagement and related terms. *Calif Sch Psychologist*. 2003;8(1):7-27. doi:10.1007/bf03340893

7. Axelson RD, Flick A. Defining student engagement. *Change Mag High Learn*. 2010;43(1):38-43. doi:10.1080/00091383.2011.533096

8. Ainley M. What do we know about student motivation and engagement? Presented at the annual meeting of the Australian Association for Research in Education, Melbourne: 2004.

9. Pickering JD. Cognitive engagement: A more reliable proxy for learning? *Medical Sci Educ*. 2017;27(4):821-823. doi:10.1007/s40670-017-0447-8

10. Blumenfeld PC, Kempler TM, Krajcik JS. The Cambridge Handbook of the Learning Sciences. 2005:475-488. doi:10.1017/cbo9780511816833.029

11. Armstrong EG, Mackey M, Spear SJ. Medical education as a process management problem. *Acad Med*. 2004;79(8):721-728. doi:10.1097/00001888-200408000-00002

12. Jones BD. Motivating students to engage in learning: The MUSIC Model of Academic Motivation. *International Journal of Teaching and Learning in Higher Education*. 2009;21(2):272-285. https://eric.ed.gov/?id=EJ899315.

13. Manning C. “Senioritis:” An Analysis of Academic Motivation and Burnout in College Students through the Lens of Positive Psychology. *Psychology Honors Papers*. 2011; 13. http://digitalcommons.conncoll.edu/psychhp/13.

14. Deci EL, Ryan RM. Self-determination theory: A macrotheory of human motivation, development, and health. *Can Psychology Psychologie Can*. 2008;49(3):182-185. doi:10.1037/a0012801

15. Maslow AH. A theory of human motivation. *Psychol Rev*. 1943;50(4):370. doi:10.1037/h0054346

16. Keller JM, Dodge B, Keller B, Sari F, Stevens G. The ARCS (Attention, Relevant, Confidence & Satisfaction) Model of motivational strategies for course designers and developers. 1982.

17. Pelaccia T, Viau R. Motivation in medical education*. *Med Teach*. 1AD;39(2):1-5. doi:10.1080/0142159x.2016.1248924

18. Deci EL, Ryan RM. Intrinsic Motivation and Self-Determination in Human Behavior. New York, NY: Plenum Press; 1985. doi:10.1007/978-1-4899-2271-7

19. Saeed S, Zyngier D. How motivation influences student engagement: A qualitative case study. *J Educ Learn*. 2012;1(2). doi:10.5539/jel.v1n2p252

20. Nayir F. The relationship between student motivation and class engagement levels. *Eurasian J Educ Res*. 2017;17(71):59-78. doi:10.14689/ejer.2017.71.4

21. Irvin J, Meltzer J, Dukes M. Taking Action on Adolescent Literacy: An Implementation Guide for School Leaders. Virginia: Association for Supervision and Curriculum Development; 2007.

22. Kusurkar R, Ten Cate TJ, van Asperen M, Croiset G. Motivation as an independent and a dependent variable in medical education: A review of the literature. *Med Teach*. 2011;33(5):e242-e262. doi:10.3109/0142159x.2011.558539

23. Vallerand RJ, Pelletier LG, Blais MR, Briere NM, Senecal C, Vallieres EF. The Academic Motivation Scale: A measure of intrinsic, extrinsic, and amotivation in education. *Educ Psychol Meas*. 1992;52(4):1003-1017. doi:10.1177/0013164492052004025

24. Deci EL, Ryan RM. A motivational approach to self: Integration in personality. *University of Nebraska Press*. 1991:237.

25. Sobral DT. What kind of motivation drives medical students' learning quests? *Med Educ*. 2004;38(9):950-957. doi:10.1111/j.1365-2929.2004.01913.x

26. Park J, Chung S, An H, et al. A structural model of stress, motivation, and academic performance in medical students. *Psychiatry Investig*. 2012;9(2):143-149. doi:10.4306/pi.2012.9.2.143

27. Eccles JS, Wigfield A. In the mind of the actor: The structure of adolescents' achievement task values and expectancy-related beliefs. *Pers Soc Psychol B*. 1995;21(3):215-225. doi:10.1177/0146167295213003

28. Wigfield A, Eccles JS. Expectancy–Value Theory of achievement motivation. *Contemp Educ Psychol*. 2000;25(1):68-81. doi:10.1006/ceps.1999.1015

29. Lens W, Simons J, Dewitte S. Student motivation and self-regulation as a function of future time perspective and perceived instrumentality. 2001.

30. Simons J, Dewitte S, Lens W. The role of different types of instrumentality in motivation, study strategies, and performance: Know why you learn, so you'll know what you learn! *Brit J Educ Psychol*. 2004;74(3):343-360. doi:10.1348/0007099041552314

31. Simons J, Vansteenkiste M, Lens W, Lacante M. Placing motivation and future time perspective theory in a temporal perspective. *Educ Psychol Rev*. 2004;16(2):121-139. doi:10.1023/b:edpr.0000026609.94841.2f

32. Bandura A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychol Rev*. 1977;84(2):191. doi:10.1037/0033-295x.84.2.191

33. Maraghi M, Mortazavi-Tabatabaei SA, Ahmady S, Hosseini M. The relation of educational self-efficacy and motivation among Medical Education students. *Journal of Advances in Medical Education*. 2018;1(2):1-5.

34. Malkoç A, Mutlu AK. Academic self-efficacy and academic procrastination: Exploring the mediating role of academic motivation in Turkish university students. *Univers J Educ Res*. 2018;6(10):2087-2093. doi:10.13189/ujer.2018.061005

35. Marsh HW. A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification. *Educ Psychol Rev*. 1990;2(2):77-172. doi:10.1007/bf01322177

36. Csikszentmihalyi M, Rathunde K. The measurement of flow in everyday life: toward a theory of emergent motivation. *Neb Symposium Motivation Neb Symposium Motivation*. 1992;40:57-97.

37. Rheinberg F. Intrinsic Motivation and Flow. In: Heckhausen J, Heckhausen H, eds. Motivation and Action. Cambridge: Cambridge University Press; 2008:323-348. doi:10.1017/cbo9780511499821.014

38. Baumeister RF, Leary MR. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*. 1995;117(3):497-529. doi:10.1037/0033-2909.117.3.497

39. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist*. 2000;55.

40. Jones BD, Skaggs G. Measuring students' motivation: Validity evidence for the MUSIC Model of Academic Motivation Inventory. *Int J Scholarsh Teach Learn*. 2016;10(1). doi:10.20429/ijsotl.2016.100107

41. Jones BD. User guide for assessing the components of the MUSIC Model of Academic Motivation. 2017.

42. Jones BD, Wilkins JLM. Testing the MUSIC Model of Academic Motivation through confirmatory factor analysis. *Educ Psychol-uk*. 2013:1-22. doi:10.1080/01443410.2013.785044

43. Schram AB. *Validating an Icelandic version of the MUSIC Model of Academic Motivation Inventory* [dissertation]. Blacksburg, VA: Virginia Polytechnic Institute and State University; 2015.

44. Jones BD, Li M, Cruz JM. A cross-cultural validation of the MUSIC® Model of Academic Motivation Inventory: Evidence from Chinese- and Spanish-speaking university students. *Int J Educ Psychology*. 2017;6(1):25-44. doi:10.17583/ijep.2017.2357

45. Jones BD, Byrnes MK, Jones MW. Validation of the MUSIC Model of Academic Motivation Inventory: Evidence for use with veterinary medicine students. *Frontiers Vet Sci*. 2019;6:11. doi:10.3389/fvets.2019.00011

46. Pace AC, Ham A-JL, Poole TM, Wahaib KL. Validation of the MUSIC® Model of Academic Motivation Inventory for use with student pharmacists. *Curr Pharm Teach Learn*. 2016;8(5):589-597. doi:10.1016/j.cptl.2016.06.001

47. Tendhar C. *Effects of Motivational Beliefs and Instructional Practice on Students' Intention to Pursue Majors and Careers in Engineering* [dissertation]. Blacksburg, VA: Virginia Polytechnic Institute and State University; 2015.

48. Papp I, Markkanen M, Bonsdorff M von. Clinical environment as a learning environment: student nurses' perceptions concerning clinical learning experiences. *Nurs Educ Today*. 2003;23(4):262-268. doi:10.1016/s0260-6917(02)00185-5

49. RStudio Team. *RStudio: Integrated Development for R* [computer software]. RStudio, Inc., Boston, MA URL http://www.rstudio.com/; 2019.

50. Linacre JM. *Winsteps® Rasch measurement computer program* [computer software]. Beaverton, Oregon: Winsteps.com; 2019.

51. Revelle W. 2018. *psych: Procedures for personality and psychological research*, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.4.

52. Dziuban CD, Shirkey EC. When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychol Bull*. 1974;81(6):358. doi:10.1037/h0036316

53. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334. doi:10.1007/bf02310555

54. Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Medical Educ*. 2011;2:53-55. doi:10.5116/ijme.4dfb.8dfd

55. Yves R. lavaan: An R Package for Structural Equation Modeling. *J Stat Softw*. 2012;48(2):68-81. doi: 10.18637/jss.v048.i02.

56. Granger C. Rasch Analysis is important to understand and use for measurement. *Rasch Measurement Transactions*. 2008;21(3):1122-1123. https://www.rasch.org/rmt/rmt213d.htm.

57. Boone WJ, Noltemeyer A. Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Educ*. 2017;4(1). doi:10.1080/2331186x.2017.1416898

58. Duncan PW, Bode RK, Lai SM, Perera S, Investigators GA in NA. Rasch analysis of a new stroke-specific outcome scale: the stroke impact. *Arch Phys Med Rehab*. 2003;84(7):950-963. doi:10.1016/s0003-9993(03)00035-2

59. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthrit Care Res*. 2007;57(8):1358-1362. doi:10.1002/art.23108

60. Andrich D. A rating formulation for ordered response categories. *Psychometrika*. 1978;43(4):561-573. doi:10.1007/bf02293814

61. Andersen EB. The rating scale model. In: van der Linden WJ, Hambleton RK, eds., *Handbook of Modern Item Response Theory*. New York: Springer Science+Business Media; 1997:67-84. doi:10.1007/978-1-4757-2691-6_4

62. Wright BD. Model selection: Rating Scale Model (RSM) or Partial Credit Model (PCM)? *Rasch Measurement Transactions*. 1998;12(3):641-642.

63. Linacre JM. Comparing and choosing between "Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). *Rasch Measurement Transactions*. 2000;14(3):768.

64. Shi Y, Gugiu PC, Crowe RP, Way DP. A Rasch analysis validation of the Maslach Burnout Inventory–Student Survey with preclinical medical students. *Teach Learn Med*. 2018;31(2):1-16. doi:10.1080/10401334.2018.1523010

65. Wright B, Stone M. *Measurement Essentials*. 2nd Edition. Wilmington, Delaware: Wide Range, Inc.; 1999.

66. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Struct Equ Model Multidiscip J*. 1999;6(1):1-55. doi:10.1080/10705519909540118

67. Browne MW, Cudeck R. Alternative ways of assessing model fit. *Sociol Method Res*. 1992;21(2):230-258. doi:10.1177/0049124192021002005

68. Linacre JM. What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions*. 2002;16(2):878.

69. Hinkle DE, Wiersma W, Jurs SG. Applied Statistics for the Behavioral Sciences. *Houghton Mifflin*. 2001. https://books.google.co.nz/books?id=7tntAAAAMAAJ.

70. New Zealand MSOD Steering Group. National report on students commencing medical school in New Zealand in 2014. https://www.otago.ac.nz/medicine/about/accountability/amc-accreditation/otago719562.pdf?ssSourceSiteId=medical-school. Updated September 2017. Accessed January 17, 2019.

71. New Zealand MSOD Steering Group. National report on students commencing medical school in New Zealand in 2015-2019. https://www.otago.ac.nz/medicine/about/accountability/amc-accreditation/otago718595.pdf?ssSourceSiteId=medical-school. Updated July 2019. Accessed January 17, 2019.

Table 1. the MUSIC® Inventory constructs and their definitions.

MUSIC Model Constructs	Definitions The degree to which a student perceives that:	Related constructs
Empowerment	he or she has control of his or her learning environment in the course	Autonomy (Deci & Ryan, 1991)
Usefulness	the coursework is useful to his or her future	Utility value (Wigfield & Eccles, 2000)
Success	he or she can succeed at the coursework	Expectancy for success (Wigfield & Eccles, 2000)
Interest	the instructional methods and coursework are interesting	Situational interest (Hidi & Renninger, 2006)
Caring	the instructor cares about whether the student succeeds in the coursework and cares about the student's well-being	Caring (Noddings, 1992)

From the "User Guide for Assessing the Components of the MUSIC® Model of Motivation" (2017) by B. D. Jones available at www.theMUSICmodel.com © Brett D. Jones.

Table 2: Demographic data of sample compared with population

Demographics	Category	Number	Percent	ALM Population extrapolated from 2014-2016 entrance data
Age[1 2]	Under 21	3	1.9%	
	21-24	108	69.7%	62.4%
	25-35	39	25.2%	34.3%
	36-60	5	3.2%	3.3%
Gender[1 2]	Male	61	39.4%	42.2%
	Female	92	59.4%	57.8%
	Prefer not to say	2	1.3%	--
Campus[3]	Wellington	62	40.0%	36.5%
	Christchurch	43	27.7%	36.8%
	Dunedin	50	32.2%	26.6%
Year[3]	4th	64	41.3%	33.8%
	5th	38	24.5%	32.1%
	6th	53	34.2%	34.1%
Ethnicity[1 2]	NZ European	109	70.3%	57.2%
	Maori	12	7.7%	13.1%
	Pacific	3	1.9%	5.9%
	Other	30	19.4%	40.6%
Mode of entry	First Year	109	70.3%	
	Graduate	30	19.4%	
	Alternate	16	10.3%	

Note: [1]Population data extrapolated from: New Zealand MSOD Steering Group (September 2017). National report on students commencing medical school in New Zealand in 2014.; [2]New Zealand MSOD Steering Group (July 2019). National report on students commencing medical school in New Zealand in 2015-2019; and [3]course enrolments in the university learning management system.

Table 3. Cronbach's alpha and corrected Pearson's correlation coefficients.

	Empowerment	Usefulness	Success	Interest	Caring
Empowerment	**.87**	.65	.51	.69	.52
Usefulness		**.92**	.38	.89	.49
Success			**.90**	.44	.57
Interest				**.91**	.58
Caring					**.89**

Note: inter-item correlation for each scale (Cronbach's alpha) in bold on the diagonal; corrected Pearson's correlation coefficients between scales above the diagonal

Table 4. Factor loadings for five-factor solutions.

	Empowerment		Usefulness		Success		Interest		Caring	
	Full	Edited	Full	Edited	Full	Edited	Full	Edited	Full	Edited
Empower 1	0.75	0.76								
Empower 2	0.83	0.82								
Empower 3	0.69	0.72								
Empower4	0.57	0.58								
Empower5	0.68	0.68								
Useful 1			0.65	0.58						
Useful 2			0.73	0.66						
Useful 3			0.82	0.81						
Useful 4			0.70	0.68						
Useful 5			0.79	0.78						
Success 1					0.83	0.82				
Success 2					0.80	0.79				
Success 3					0.77	0.78				
Success 4					0.89	0.90				
Interest 1			0.41							
Interest 2							0.84	0.85		
Interest 3							0.60	0.74		
Interest 4							0.48	0.59		
Interest 5			0.53							
Interest 6			0.63							
Caring 1									0.81	0.81
Caring 2									0.86	0.86
Caring 3									0.57	0.55
Caring 4									0.59	0.59
Caring 5									0.76	0.76
Caring 6									0.63	0.62

Note: Column one (“full”) shows the factor loadings using the full set of items. Column two (“edited”) shows the loadings with the three Interest items related to coursework removed.

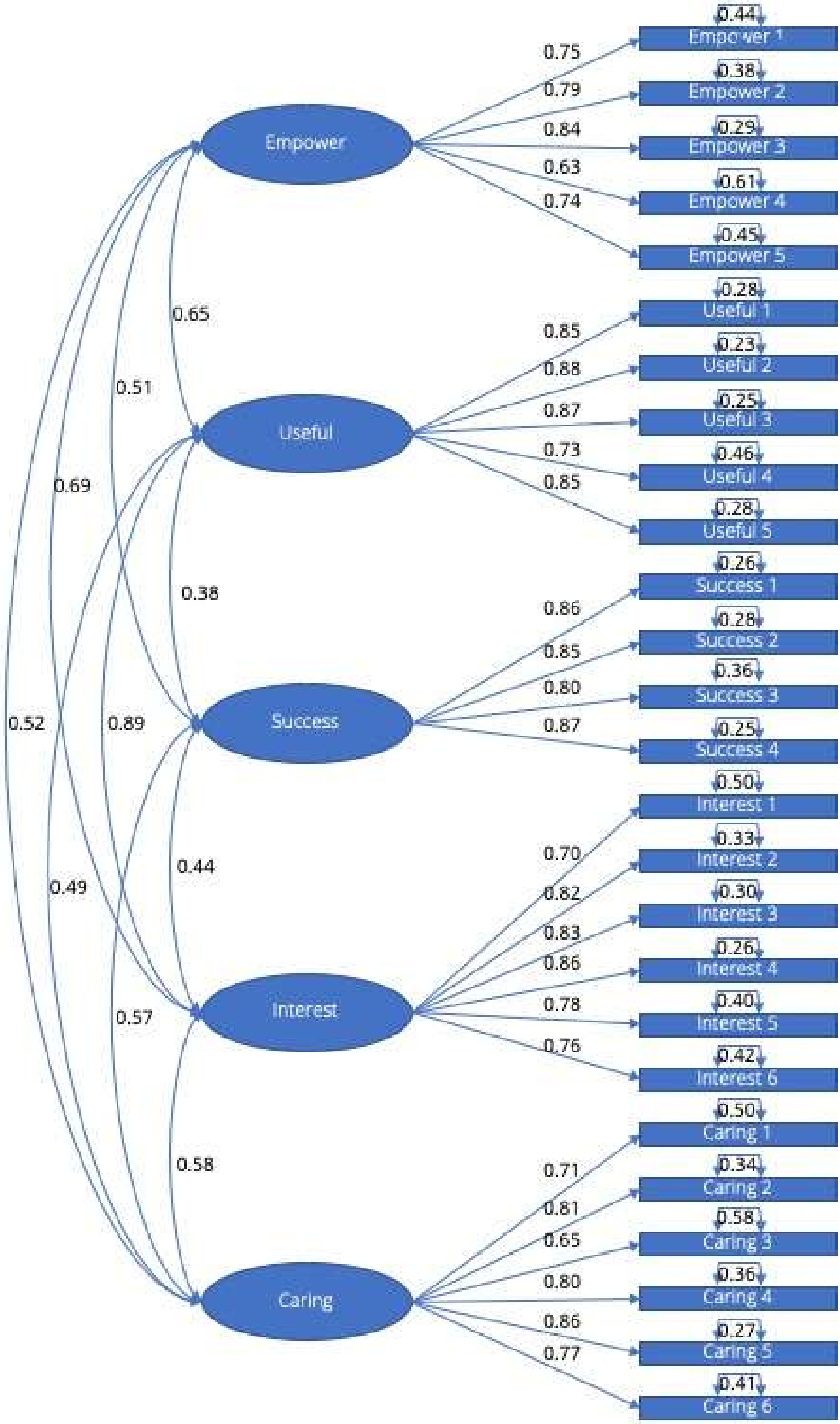
Table 5. Rasch analysis item fit statistics, item difficulties, point-measure correlations and scale person and item separation and reliability.

Scale Item	Item Difficulty	Model S.E.	Infit MSQ	Outfit MSQ	Point-measure correlation
Empowerment					
Empower 5	.40	.11	0.89	0.87	.79
Empower 3	.27	.11	0.75	0.73	.84
Empower 2	.14	.11	0.94	0.93	.83
Empower 1	-.11	.11	1.18	1.12	.80
Empower 4	-.70	.11	1.24	1.17	.73
Usefulness					
Useful 3	.65	.15	0.79	0.77	.88
Useful 1	.26	.15	1.06	1.09	.82
Useful 2	-.22	.16	0.93	0.83	.84
Useful 4	-.35	.16	1.25	1.13	.78
Useful 5	-.35	.16	0.83	0.82	.84
Success					
Success 3	.89	.15	1.17	1.06	.86
Success 4	.24	.16	0.80	0.76	.84
Success 1	-.50	.17	1.00	0.97	.85
Success 2	-.62	.17	0.92	0.84	.83
Interest					
Interest 5	.77	.13	1.18	1.16	.81
Interest 4	.72	.13	0.72	0.74	.86
Interest 2	.70	.13	0.81	0.80	.84
Interest 3	.63	.13	0.93	0.90	.85
Interest 6	-1.40	.14	1.15	1.09	.79
Interest 1	-1.42	.14	1.23	1.32	.75

Caring					
Caring 6	1.39	.12	0.99	0.99	.83
Caring 3	1.24	.12	1.21	1.31	.72
Caring 1	-.17	.14	1.37	1.30	.77
Caring 2	-.47	.14	0.87	0.84	.81
Caring 5	-.70	.15	0.70	0.65	.82
Caring 4	-1.28	.15	0.91	0.80	.77

		Person		Item	
		Separation	Reliability	Separation	Reliability
Empowerment					
	Real	2.28	.84	3.27	.91
	Model	2.62	.87	3.43	.92
Usefulness					
	Real	2.85	.89	2.25	.83
	Model	3.19	.91	2.33	.84
Success					
	Real	2.63	.87	3.48	.92
	Model	3.03	.90	3.55	.93
Interest					
	Real	2.67	.88	6.97	.98
	Model	3.09	.91	7.33	.98
Caring					
	Real	2.42	.85	6.73	.98
	Model	2.78	.89	7.04	.98

Figure 1: Path diagram of theoretically based five-factor model of MUSIC® Inventory.



Appendix 1: MUSIC Inventory (Medical Student version, present tense, program level)

Instructions

Please rate the items below using the following scale:

1 Strongly Disagree	2 Disagree	3 Somewhat disagree	4 Somewhat agree	5 Agree	6 Strongly agree

Thinking about the medical school programme and the courses you have taken and are currently taking, please rate your level of agreement or disagreement with the following statements:

____ 1. The coursework holds my attention
Interest

____ 2. I have the opportunity to decide for myself how to meet course goals.
Empowerment

____ 3. In general, the coursework is useful to me.
Usefulness

____ 4. The tutors are available to answer my questions about the coursework. Caring

____ 5. I am confident that I can succeed in the coursework. Success

Appendix 2: Rasch analysis by scale

Empowerment

To ensure that each factor was unidimensional, a test of dimensionality was performed. For the empowerment factor no contrast was found to measure over 2.0, suggesting that the factor is unidimensional.[39] Table 5 shows the item difficulties (0.40 – -0.70) from most difficult (highest score) to easiest (lowest score) for the items that make up the Empowerment scale of the MUSIC® Inventory. Item difficulties can be thought of as the ease with which a respondent will endorse the item. For example, medical students found the item, “I have flexibility in what I am allowed to do in the courses.” the most difficult to endorse or agree with, while the item, “I have control over how I learn the course content.” was the easiest to endorse or agree with. Empowerment showed excellent item separation and reliability, with item separation of 3.27, item reliability of .91. Person separation and reliability were relatively lower, though still good (separation = 2.28, reliability = .84).

Person-item maps (Wright Maps) graphically present the distribution of person responses to items on the left side of the map, and the distribution of item difficulties on the right side of the map. The mean person distribution is represented by the M on the left of the figure, and the mean for the item difficulties is represented by the M on the right of the figure. The Wright Map for Empowerment is presented in Figure 2. The mean of each side of the graph is less than 0.5 logits apart indicating that the two distributions are reasonably well aligned.[53] However, the five items of empowerment cluster within a two-logit span around the mean difficulty. This indicates that while there is some difference in the ability to endorse, or agree with, the items, they are very similar. This similarity indicates some redundancy in the items measuring empowerment.

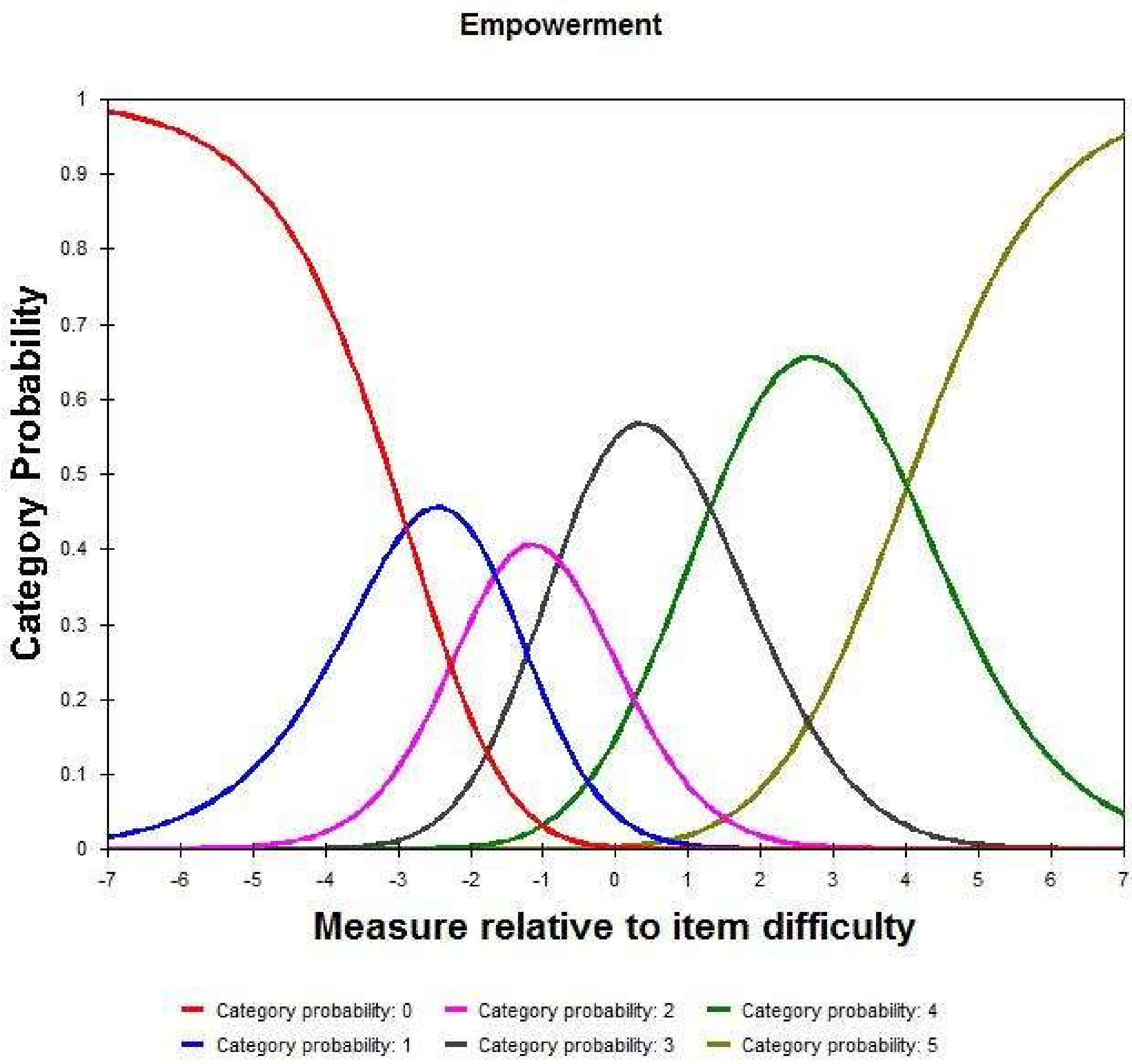
Figure 2: Wright map showing the person-item relationship for the items measuring empowerment

```
-----------------------------------------------------------

MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    4                +
                  .  |
                     |
                    T|
               .###  |
                     |
    3                +
                     |
              #####  |
                     |
                     |
             .####  |
    2                +
                    S|
              #####  |
                     |
                     |
               .###  |
    1                +
               ####  |T
                     |
                     |
         .#########  |S Empower3  Empower5
                    M|  Empower2
    0       .######  +M
                     |  Empower1
               ####  |S
                     |
              .####  |  Empower4
             .#####  |T
   -1                +
               .###  |
               ####  |
                     |
                 ## S|
               .###  |
   -2                +
                  .  |
                     |
                 .#  |
                     |
                  #  |
   -3                +
                 .#  |
                    T|
                     |
                     |
                     |
   -4                +
                     |
                     |
                 .#  |
                     |
                     |
   -5             #  +
               <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

Figure 3 shows the probability curves for each response category for the empowerment scale. From this graph, each category shows a clear area, or threshold, along the continuum where it is more likely to be chosen. This confirms the Likert-type items are differentiating people at each category threshold.

Figure 3: Category probability curves for the empowerment scale.



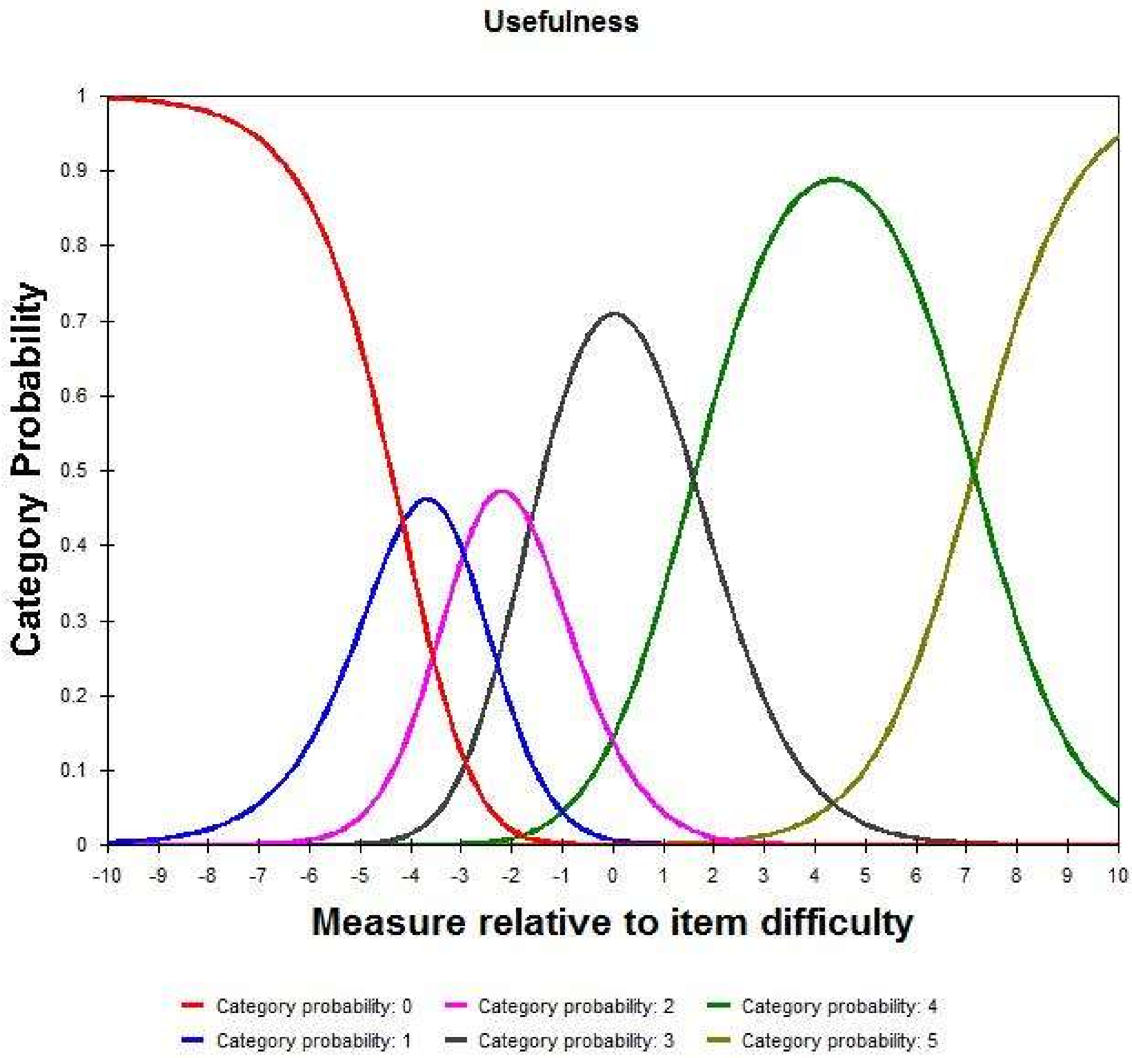
Usefulness

The dimensionality test indicated that the factor is unidimensional with no contrast over 2.0. Table 5 shows the items difficulties for each of the five items (range: 0.65 – -0.35). Similar to the empowerment scale, the Wright map for usefulness (Figure 4) shows the items clumped around the mean, distributed across slightly more than one logit. The inference from this is that the items are measuring similar aspects of the usefulness factor. There is also possible redundancy in the items as seen by items 2, 4, and 5 located at the same location along the distribution. The mean distribution of the persons is over two logits higher than that for items. The higher mean for the person distribution suggests that students found it relatively easy to endorse the useful items, pointing to a possible ceiling effect. No category peak in the category probably curves (Figure 5) is buried in another, indicating the respondents' ability to distinguish between the options for the items.

Figure 4: Wright map for the usefulness scale

```
---------------------------------------------------------------
MEASURE    PERSON - MAP - ITEM
                  <more>|<rare>
    9             #  +
                     |
                  # T|
                     |
    8                +
                     |
                .##  |
                     |
    7                +
                 ##  |
                     |
                     |
    6                +
              .####  |
                    S|
                     |
    5                +
                     |
                     |
         #########   |
    4                +
                     |
                     |
                     |
    3         .####  +
                     |
                     |
                    M|
    2         .####  +
                     |
                     |
               .###  |
    1                +
               .###  |T Useful3
                     |S
                     |  Useful1
    0           ###  +M
                     |  Useful2  Useful4  Useful5
                .##  |S
                    S|T
   -1            .#  +
                     |
                .##  |
                  .  |
   -2                +
                 .#  |
                  .  |
                     |
   -3             .  +
                     |
                     |
                  .  |
   -4               T+
                     |
                  .  |
                     |
   -5             #  +
                  <less>|<freq>
 EACH "#" IS 3: EACH "." IS 1 TO 2
```

Figure 5: Category probability curves for the usefulness scale



Success

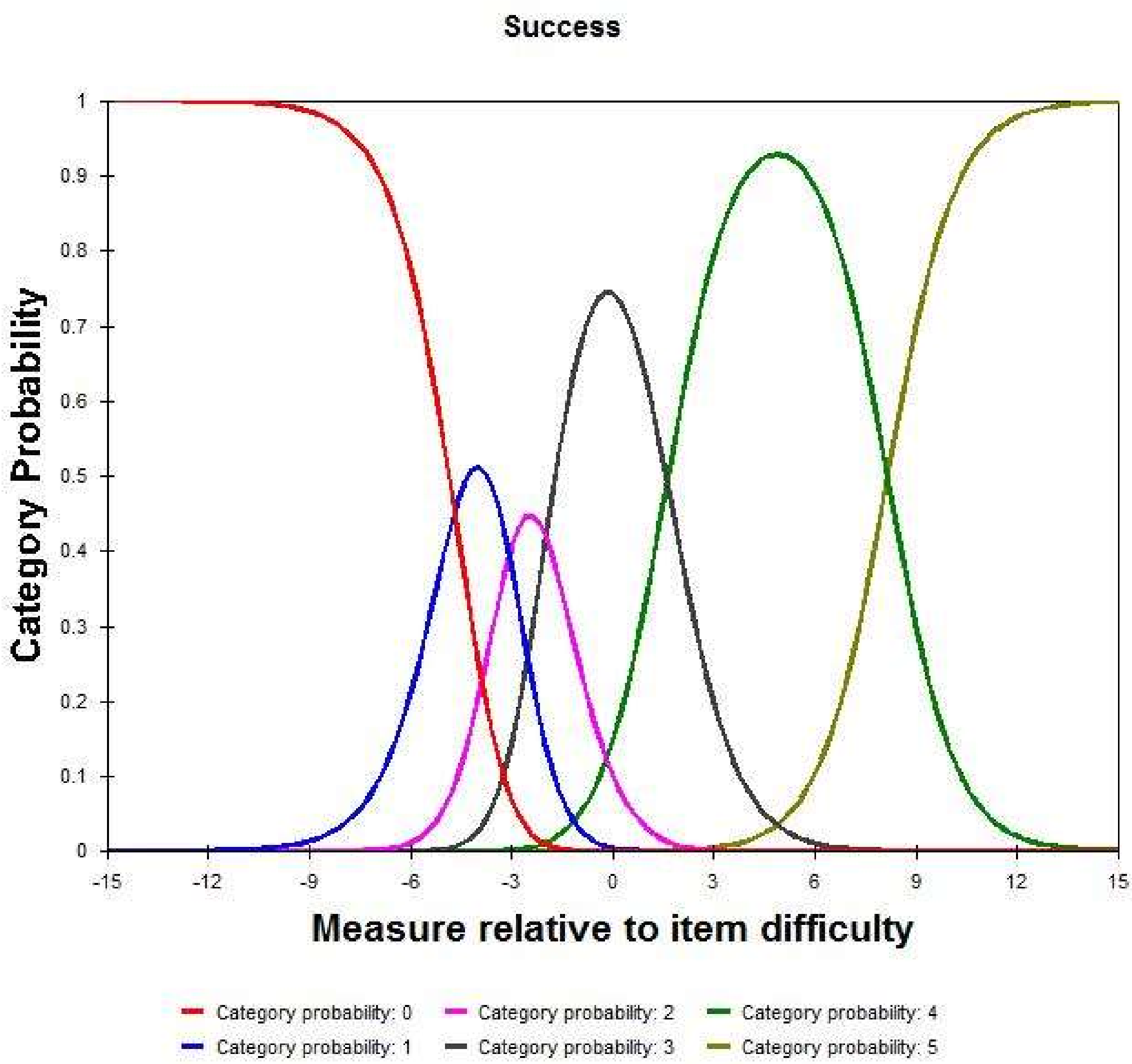
The success scale showed no contrasts greater than 2.0 in the test of dimensionality, indicating a unidimensional measure. Item difficulties are shown in Table 5 and range from 0.89 to -0.62. Person and item separation and reliability indicated a good fit to the model (Table 5). Similar to the Wright-map for the Usefulness scale, the means for persons and items were greater than 2.0 logits indicating a slight ceiling effect for the person distribution (Figure 6). The four success items are clustered within a 2-logit distribution at the midpoint of the logit scale and there is the possibility of redundancy with items 1 and 2.

Figure 6: Wright map for the success scale

```
MEASURE    PERSON - MAP - ITEM
                   <more>|<rare>
   10           .###  +
                      |
                  ##  |
    9                T+
                      |
                      |
    8              .#  +
                      |
                      |
    7             .##  +
                      |
                      |
    6                 +
                     S|
                      |
    5      .#########  +
                      |
                      |
    4                 +
                      |
                      |
    3                 +
          .######     |
                     M|
    2                 +
            ######    |
                      |T
    1                 +  Success3
              ####    |S
                      |  Success4
    0          .####  +M
                      |
                      |S Success1  Success2
   -1             ### S+
                   .#  |T
                      |
   -2              .#  +
                   .#  |
                    .  |
   -3                 +
                    .  |
                    .  |
   -4               .  +
                     T|
                    .  |
   -5                 +
                      |
                      |
   -6               .  +
                   <less>|<freq>
 EACH "#" IS 3: EACH "." IS 1 TO 2
```

Figure 7 shows the probability curves for the success scale. Each category has a clear peak with no buried categories indicating good differentiation of scores across the response categories.

Figure 7: Category probability curves for the success scale



Interest

When testing the Interest items for dimensionality, one contrast showed an eigenvalue of 2.15, indicating a possible second dimension in the item set. Items 2–4 were located within cluster one of contrast one and items 1, 5, and 6 were located within cluster three of contrast one. This clustering is similar to the factor loading differences found in the confirmatory factor analysis. However, the model fit was good for the scale, with item separation and reliability of 6.97 and 0.98 respectively, and person separation and reliability of 2.67 and 0.88 respectively (Table 5).

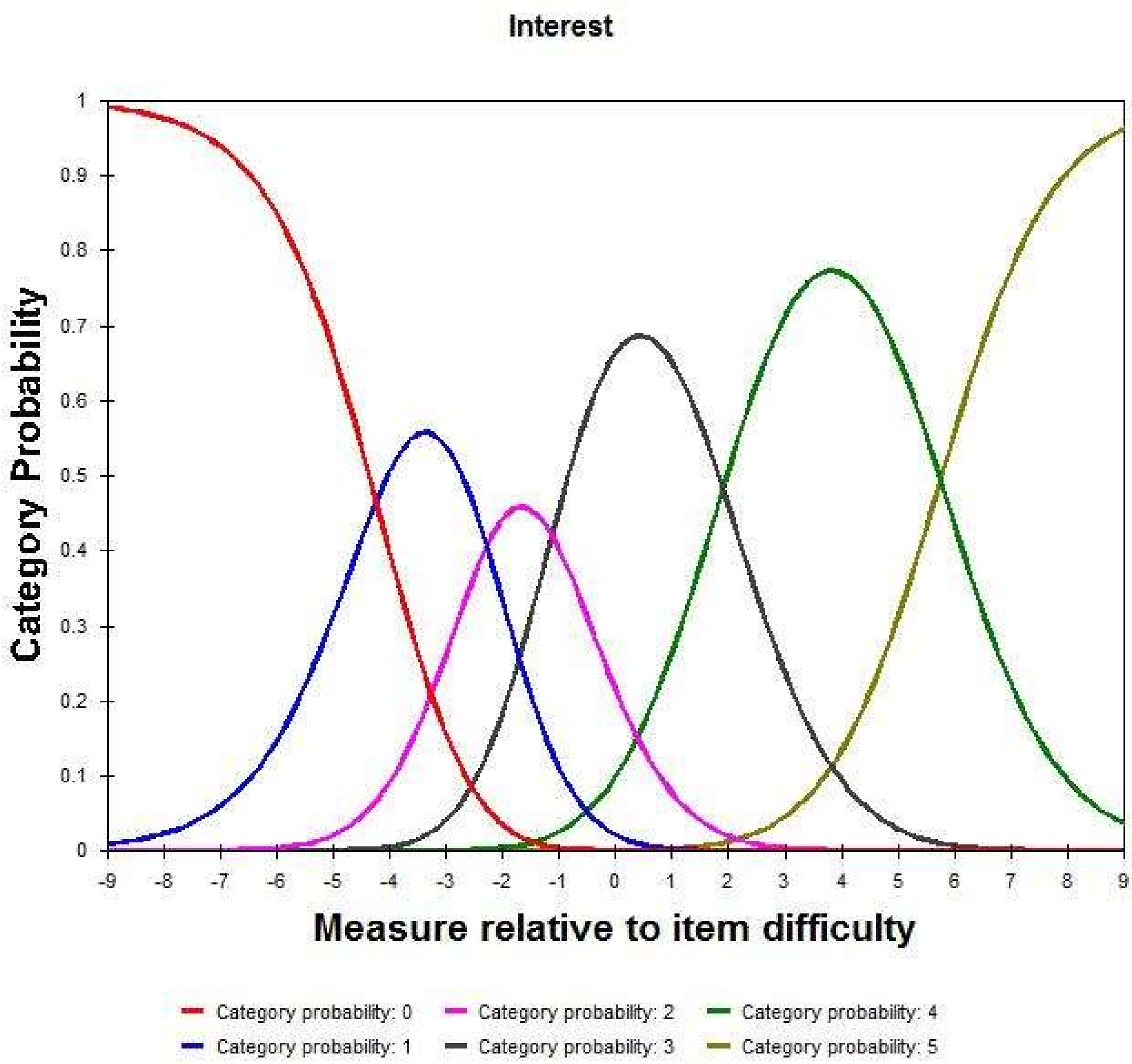
Scrutiny of the Wright Map shows similar distributions of person and item measures with the average person and item option measures less than 1 logit apart. However, when looking at the item options, items 1 ("The coursework holds my attention") and 6 ("The coursework is interesting to me") of the interest scale appear to be consistently easier to endorse than items 2–5 (Figure 8).

Figure 8: Wright map for the interest scale

```
MEASURE    PERSON - MAP - ITEM
                <more>|<rare>
    6              #  +
                      |
                      |
                   # T|
    5                 +
                      |
                  .#  |
                      |
    4                 +
                ####  |
                      |
              .#####  |
    3                S+
                      |
             #######  |
                      |
    2          #####  +T
                      |
            .#######  |
                      |
    1      .########  +S
                     M|   Interest2  Interest3  Interest4  Interest5
            .#######  |
                      |
    0          #####  +M
                 .##  |
                      |
              .#####  |
   -1             .#  +S
                   #  |
                  .# S|   Interest1  Interest6
                      |
   -2             .#  +T
                  .#  |
                   #  |
                      |
   -3                 +
                 .##  |
                   # T|
                   #  |
   -4                 +
                   .  |
                      |
                      |
   -5                 +
                      |
                   #  |
                      |
   -6              .  +
                <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

The items also line up together on two points in the distribution, with items 2–5 approximately one logit above the mean, and items 1 and 6 nearly two logits below the mean. From the map, the items appear to be redundantly measuring two aspects of interest. The category probability curves (Figure 9) show no disordered thresholds for the categories, with clear peaks for each category.

Figure 9: Category probability curves for the interest scale



Caring

For the caring scale, the dimensionality test confirmed a unidimensional measure with no contrast greater that 2.0. Table 5 shows the item difficulties from highest to lowest and the person and item separation and reliability. In spite of the good personal and item separations, the Wright map (Figure 10) indicates a large difference between the mean person and item option distributions (>1.5 logits).

Figure 10: Wright map for the caring scale

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
    8                +
                     |
                     |
                 .#  |
    7                +
                     |
                     |
                    T|
    6             #  +
                     |
                     |
                ###  |
    5                +
                .##  |
                     |
                     |
    4        ###### S+
                     |
                     |
           .#######  |
    3                +
           .#######  |
                     |
                     |
    2    ##########  +T
                    M|
           .#######  |  Caring6
                     |  Caring3
    1       #######  +S
               .###  |
                     |
               ####  |
    0            ##  +M
              .#### S|  Caring1
                     |  Caring2
                 .#  |  Caring5
   -1            ##  +S
                  .  |  Caring4
                  #  |
                  .  |
   -2             .  +T
                     |
                    T|
                  .  |
   -3             .  +
                     |
                     |
                     |
   -4             #  +
                     |
                     |
                     |
   -5                +
                     |
                  .  |
                     |
   -6                +
               <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

Overall, students seemed to find the caring items easier to endorse as shown by the higher mean value indicating a slight ceiling effect. Of all the scales, the items in the caring scale show the broadest distribution, though there is a significant gap between items 3 and 6 at the more difficult end of the map, and items 1, 2, 4, and 5 at the easier to endorse end of the map. There is no redundancy in the items, but the gap indicates the possibility that there is an opportunity to improve the scale with the thoughtful addition of items to fill the space between the easy to endorse and more difficult to endorse items. Figure 11 shows the probability curves for the caring scale. Each category has a clear peak with no buried categories indicating good differentiation of scores across the response categories.

Figure 11: Category probability curves for the caring scale